Isoprenyl diphosphate synthases: Protein sequence comparisons, a phylogenetic tree, and predictions of secondary structure

A. CHEN, P. A. KROON and C. D. POULTER

Protein Sci. 1994 3: 600-607

Protein Science (1994), *3*:600-607. Cambridge University Press. Printed in the USA.

Isoprenyl diphosphate synthases: Protein sequence comparisons, a phylogenetic tree, and predictions of secondary structure

ANJUN CHEN,[1] PAULOS A. KROON,[2] AND C. DALE POULTER[1]

[1] Department of Chemistry, University of Utah, Salt Lake City, Utah 84112
[2] Department of Biochemistry, University of Queensland, Herston, Queensland 4029, Australia

(RECEIVED November 2, 1993; ACCEPTED February 2, 1994)

Abstract

Isoprenyl diphosphate synthases are ubiquitous enzymes that catalyze the basic chain-elongation reaction in the isoprene biosynthetic pathway. Pairwise sequence comparisons were made for 6 farnesyl diphosphate synthases, 6 geranylgeranyl diphosphate synthases, and a hexaprenyl diphosphate synthase. Five regions with highly conserved residues, two of which contain aspartate-rich DDXX(XX)D motifs found in many prenyltransferases, were identified. A consensus secondary structure for the group, consisting mostly of α-helices, was predicted for the multiply aligned sequences from amino acid compositions, computer assignments of local structure, and hydropathy indices. Progressive sequence alignments suggest that the 13 isoprenyl diphosphate synthases evolved from a common ancestor into 3 distinct clusters. The most distant separation is between yeast hexaprenyl diphosphate synthetase and the other enzymes. Except for the chromoplastic geranylgeranyl diphosphate synthase from *Capsicum annuum*, the remaining farnesyl and geranylgeranyl diphosphate synthases segregate into prokaryotic/archaebacterial and eukaryotic families.

Keywords: catalytic site; evolution; farnesyl diphosphate; geranylgeranyl diphosphate; prenyltransferase; secondary structure; substrate binding

With more than 23,000 known members, isoprenoids constitute the most chemically diverse family of naturally occurring compounds. Some of the more important products of the pathway are the sterols (Poulter & Rilling, 1981a), ubiquinones (Ashby & Edwards, 1990), dolichols (Matsuoka et al., 1991), carotenoids (Spurgeon & Porter, 1981), prenylated proteins (Clarke, 1992), and plant mono-, sesqui-, and diterpenes (Cane, 1981; Croteau, 1981; West, 1981). All of these compounds are derived from linear isoprenoid diphosphates synthesized from isopentenyl diphosphate and dimethylallyl diphosphate by a family of prenyltransferases that catalyze sequential condensations of IPP with allylic isoprenoid diphosphates, as shown in Figure 1. Although the chemical mechanisms of these condensation reactions are identical, the isoprenyl diphosphate synthases differ in their selectivity with respect to the chain length and double-bond stereochemistry of their respective allylic substrates and the chain length and stereochemistry of newly formed double bonds in their products (Poulter & Rilling, 1978, 1981b).

During the past few years the structural genes for several farnesyl diphosphate synthases, geranylgeranyl diphosphate synthases, and a hexaprenyl diphosphate synthase have been identified and characterized. Early sequence comparisons revealed 2 conserved DDXX(XX)D aspartate-rich domains (Ashby et al., 1990), which were thought to be binding sites for the diphosphate moieties in IPP and the allylic substrates. This proposal was supported by kinetic studies of site-directed mutants (Marrero et al., 1992). More recently, Koyama et al. (1993) identified 7 conserved regions in eubacterial and eukaryotic FPPSases, including the 2 aspartate-rich regions.

Multiple sequence alignments are valuable for identifying conserved sequences in proteins. In addition, multiple alignments can be used in conjunction with procedures for predicting secondary structure from primary sequences to obtain improved predictions, as for example, the prediction of the structure of the α subunit in tryptophan synthase (Crawford et al., 1987). We now report sequence comparisons for 13 prenyltransferases, including 6 FPPSases, 6 GGPPSases, and a HexPPSase that suggest divergence from a common ancestor based on a com-

Reprint requests to: C. Dale Poulter, Department of Chemistry, University of Utah, Salt Lake City, Utah 84112; e-mail: poulter@chemistry.utah.edu.

Abbreviations: DMAPP, dimethylallyl diphosphate; FPP, farnesyl diphosphate; FPPSase, farnesyl diphosphate synthase; GGPP, geranylgeranyl diphosphate; GGPPSase, geranylgeranyl diphosphate synthase; HexPPSase, hexaprenyl diphosphate synthase; IPP, isopentenyl diphosphate.

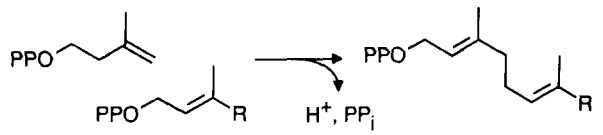

Fig. 1. Synthesis of linear isoprenoid diphosphates from IPP and DMAPP by isoprenyl diphosphate synthases.

bination of function and organisms. We also propose a common α-helical secondary structure for the 13 enzymes.

Results

Pairwise comparisons

Amino acid sequences for 6 FPPSases, 6 GGPPSases, and a HexPPSase were compared pairwise by the Needleman and Wunsch method using the TREE program of Feng and Doolittle (1987, 1990), and the results are shown in Table 1. There was 45–84% amino acid identity among the eukaryotic FPPSases from humans (Sheares et al., 1989), rats (Clarke et al., 1987), chickens (Kroon, unpubl. results), and yeast (Anderson et al., 1989). Substantially lower identities of 10–25% were seen when the eukaryotic FPPSases were compared as a group with the other 9 prenyltransferases, including the eubacterial FPPSases from *Escherichia coli* (Fujisaki et al., 1990) and *Bacillus stearothermophilus* (Koyama et al., 1993). However, the 2 eubacterial FPPSases showed substantial identities of 27–44% with the chromoplast (chloroplast-related) GGPPSase from *Capsicum annuum* (Kuntz et al., 1992), the bifunctional archaebacterial FPP/GGPPSase from *Methanobacterium thermoautotrophicum* (Chen & Poulter, unpubl. results), and the eubacterial GGPPSases from *Erwinia herbicola* (Armstrong et al., 1990; Math et al., 1992), *Erwinia uredovora* (Misawa et al., 1990), and *Rhodobacter capsulatus* (Armstrong et al., 1989). The fungal GGPPSase from *Neurospora crassa* (Carattoli et al., 1991) and yeast HexPPSase (Ashby & Edwards, 1990) had lower sequence identities of 10–25% with the other prenyltransferases.

Multiple sequence alignments, conserved sequences, and a phylogenetic tree for isoprenyl diphosphate synthase

The amino acid sequences for 13 isoprenyl diphosphates were aligned according to the procedures of Feng and Doolittle (1990), as shown in Figure 2. Five regions, designated I–V, were found where highly conserved residues appeared in at least 12 of the 13 sequences. Similar regions were identified by Koyama et al. (1993) for a more limited set of isoprenyl diphosphate synthases. Regions II and V are rich in negatively charged aspartates and positively charged arginines or lysines. These sequences correspond to those originally labeled as domains I and II, respectively, by Ashby et al. (1990), who proposed they were diphosphate binding motifs.

The relatively high pairwise percentage identities among selected pairs of prenyltransferases listed in Table 1 are consistent with all 13 isoprenyl diphosphate synthases having diverged from a common ancestral enzyme. This hypothesis is further supported by the existence of 5 highly conserved regions, 2 of which are of considerable length. However, it was apparent from inspection of the alignments that there is a high degree of divergence between the eukaryotic and eubacterial FPPSases and between the FPPSase and HexPPSase in yeast.

A phylogenetic tree (see Fig. 3) was constructed for the 13 isoprenyl diphosphate synthases using the progressive multiple alignments (Feng & Doolittle, 1987, 1990) shown in Figure 2. Three major groupings were obtained. The most primitive branch was a functional segregation of the higher chain length yeast HexPPSase from the shorter chain farnesyl and geranylgeranyl synthases. The shorter chain length enzymes further segregated into bacterial (eubacteria and archaebacteria) and eukaryotic clusters. The single exception to this pattern was the

Table 1. *Pairwise percent identity of isoprenyl diphosphate synthases*[a]

Protein[b]	FPP_ HUM	FPP_ RAT	FPP_ CHI	FPP_ YSC	FPP_ ECO	RPP_ BST	GGPP_ CAN	GGPP_ MTH	GGPP_ EHE	GGPP_ EUR	GGPP_ RCA	GGPP_ NCR	HPP_ YSC
FPP_HUM	–	84.1	68.0	45.0	23.1	22.3	17.6	21.2	18.9	18.4	18.4	12.6	16.1
FPP_RAT		–	66.0	46.0	24.8	22.7	17.6	23.0	19.5	18.4	19.2	13.1	17.0
FPP_CHI			–	46.5	22.0	24.9	18.7	23.5	18.4	18.6	19.8	12.9	16.3
FPP_YSC				–	21.2	22.6	20.7	23.9	19.4	15.7	20.2	10.3	17.0
FPP_ECO					–	42.9	34.7	30.8	31.7	31.9	29.4	15.8	23.6
FPP_BST						–	39.6	33.4	33.6	33.6	31.3	12.9	25.6
GGPP_CAN							–	28.7	27.1	27.1	25.8	12.5	20.4
GGPP_MTH								–	25.7	26.6	29.9	16.1	22.7
GGPP_EHE									–	51.8	26.4	11.8	21.8
GGPP_EUR										–	25.6	11.3	22.5
GGPP_RCA											–	13.9	20.6
GGPP_NCR												–	10.7
HPP_YSC													–

[a] Pairwise sequence comparisons were done by TREE of Feng and Doolittle (1987). The percent identity was based on the aligned regions.
[b] FPP_HUM, *Homo sapiens* FPPSase; FPP_RAT, *Rattus rattus* FPPSase; FPP_CHI, *Gallus gallus* FPPSase; FPP_YSC, yeast *Saccharomyces cerevisiae* FPPSase; FPP_ECO, *Escherichia coli* FPPSase; FPP_BST, *Bacillus stearothermophilus* FPPSase; GGPP_CAN, *Capsicum annuum* GGPPSase; GGPP_MTH, *Methanobacterium thermoautotrophicum* GGPPSase; GGPP_EHE, *Erwinia herbicola* GGPPSase; GGPP_EUR, *Erwinia uredovora* GGPPSase; GGPP_RCA, *Rhodobacter capsulatus* GGPPSase; GGPP_NCR, *Neurospora crassa* GGPPSase; HPP_YSC yeast *S. cerevisiae* HexPPSase.

Fig. 2. Multiple sequence alignment for the 13 isoprenyl diphosphate synthases listed in Table 1. Long N-terminal sequences and insertions in HPP_YSC are omitted, but the numbers of amino acids are shown in parentheses. Consensus sequences shown below the 5 highly conserved sequence domains, I–V, are double underlined. A region clearly corresponding to domain III was not seen in HPP_YSC. Residues conserved differently in eukaryotic FPP synthases are in bold. The peptide in chicken FPP synthase that was labeled during photoaffinity experiments is underlined.

inclusion of the chromoplastic GGPPSase from *C. annuum* (green peppers) in a cluster of eubacterial farnesyl and geranylgeranyl diphosphate synthases. These results indicate that the chain length selectivity of the short-chain prenyltransferases cannot be readily deduced from sequence comparisons and that assignments of function should be made biochemically.

Prediction of secondary structure

Because the pairwise alignments indicate that the isoprenyl diphosphate synthases have diverged from a common ancestor, it is reasonable to assume that the gross topological features of the ancestor were conserved during evolution. We initially compared the amino acid compositions of 11 prenyltransferases, all of the FPPSases and GGPPSases except the highly diverged *N. crassa* enzyme, using Chou's approach (Chou, 1989) for predicting structural classes of proteins from their amino acid contents. The results are shown in Table 2. As judged by comparing the average amino acid composition between the isoprenyl diphosphate synthases with those of representative all-α, all-β, $\alpha + \beta$, and α/β proteins, the prenyltransferases most closely resemble typical all-α or α/β structures, suggesting an all α-helix protein or a protein dominated by α-helices.

A consensus secondary structure for the isoprenyl diphosphate synthases was predicted from a combination of the multiple sequence alignments, probabilities for formation of loop, α-helix, and β-sheet regions (Fig. 4), and an average hydropathy plot (Fig. 5). Predictions by Garnier–Osguthorpe–Robson (GOR) or Chou–Fasman (CF) methods were in good agreement and predicted 8 α-helices and 4 short β-sheets. The location of the α-helices and β-sheets generally correlated well with the average hydropathy plot for the 11 amino acid sequences. Loops 1, 2,

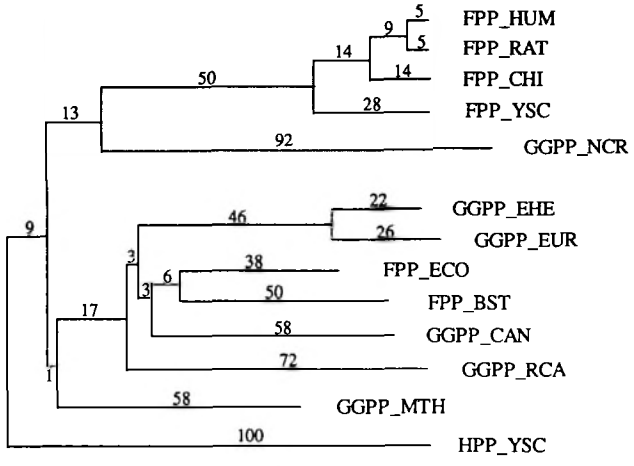

Fig. 3. A phylogenetic tree for isoprenyl diphosphate synthases constructed from progressive alignments using the TREE program and refined as described by Feng and Doolittle (1990).

Table 2. *Comparison of average amino acid compositions of isoprenyl diphosphate synthases and the 4 protein classes*[a]

Amino acid	Synthases[b]	All-α	All-β	α + β	α/β
Ala	11.2	11.6	7.3	9.3	8.3
Arg	5.1	2.2	2.4	4.1	3.4
Asn	2.5	4.0	5.0	6.4	4.2
Asp	7.0	6.7	4.4	5.9	5.6
Cys	1.6	0.9	2.7	3.9	1.5
Gln	4.6	2.7	4.4	3.9	2.6
Glu	7.3	5.5	3.1	4.6	5.9
Gly	6.7	8.1	10.7	9.1	8.7
His	2.3	4.5	1.8	1.7	2.5
Ile	5.5	3.7	4.3	4.9	5.5
Leu	11.7	9.0	6.4	5.8	7.8
Lys	6.1	10.2	4.1	5.9	7.4
Met	2.6	2.0	0.6	1.3	2.1
Phe	3.3	5.0	3.1	2.8	3.6
Pro	3.5	3.4	4.6	3.8	4.3
Ser	3.1	5.0	12.3	6.7	7.5
Thr	3.9	4.9	9.1	6.2	5.5
Trp	0.6	1.3	1.6	1.6	1.7
Tyr	3.3	2.6	4.0	5.7	3.0
Val	6.0	6.8	8.2	6.5	8.7
Difference index[c]		27.4	49.8	33.2	28.3

[a] Values for protein class all-α, all-β, α + β, and α/β are taken from Chou (1989).
[b] Average amino acid compositions of all FPP synthases and GGPP synthases (not including GGPP_NCR).
[c] Difference index = $\sum |C_{Ai} - C_{Bi}|$ as described in Methods.

3, 4, and 7 were consistently assigned by computer predicted turns, gaps in the alignments, and hydrophilic peaks in the hydropathy plot. A short β-sheet was predicted within loop 3 and within 7 by GOR and CF algorithms. However, the hydropathy plots placed these "sheet" sequences in hydrophilic regions and cast doubt on their existence. The assignments for loops 5 and 6 were based on large gaps that occurred in these regions and large negative hydropathy indices. The assignment for loop 8 was based on large negative hydropathy indices in that region. There were also gaps in the alignment between α4 and α5; however, this region was not hydrophilic, and a turn motif was not predicted by GOR or CF. Perhaps these 2 helices are joined by a spacer of variable length or are fused into a single α-helix. In addition, no turns or loops were predicted between β2 and α6 or α7 and β3. Because the average amino acid composition predicted a structure primarily composed of α-helices, the short β2 and β3 regions may be helical extensions of α6 and α7, respectively, rather than β-sheets.

A predicted average secondary structure for the isoprenyl diphosphate synthases is presented in Figure 6. The high α-helix content in the structure is consistent with the statistical prediction based on amino acid composition. The secondary structural elements were arranged to place the 5 regions containing highly conserved sequences together on the same face of the structure. Although the 3-dimensional fold is not known for any prenyltransferase, one might imagine an antiparallel orientation of α2, α3, and α4/α5 that allows loops 3, 5, and 7 to be brought together. Additional support for this folding pattern is discussed in the next section. Because the consensus structure was constructed from homologous core sequences, individual enzymes may contain some additional elements of secondary structure that lay outside of the predicted consensus regions. Likewise, the lengths of some secondary structural elements, loops, and spacers undoubtedly vary from protein to protein.

A model for substrate binding

The predicted consensus structure, along with other information about catalytic site residues, can serve as a guide for locating putative binding sites for the substrates. The 5 highly conserved regions identified in Figure 2 are located in the secondary structure as follows (see Figs. 4, 6): Region I, from loop 1 to the N-terminal part of β1; Region II, from the C-terminal half of α2 to the N-terminal half of loop 3; Region III, α5; Region IV, the C-terminus of loop 5; and Region V, from the C-terminus of α7 through loop 7. Photoaffinity experiments with an azido analog of IPP (Brems et al., 1981) labeled several amino acids from positions 157 to 188 in L5 of avian FPPSase, suggesting that this part of the enzyme interacts with the hydrophobic isopentenyl moiety in IPP. Recently, Blanchard and Karst (1993) discovered that a mutation at K197 near the C-terminus of L5 in yeast FPPSase both reduced the activity and altered chain length selectivity of the enzyme. K197 is located just beyond the region labeled in the photoaffinity studies with avian protein. These results suggest that much of L5 forms an integral part of the IPP pocket with the C-terminal end of the loop extending to the binding site for the allylic substrate.

The highly conserved DDXX(XX)D motifs in L3 and L7, as well as the arginine doublet in L3, are likely candidates for diphosphate binding sites. These predictions (Ashby et al., 1990) are consistent with site-directed mutagenesis experiments (Joly & Edwards, 1993; Song & Poulter, 1994) that established that all of these residues except the last aspartate in L7 were essential for catalysis. Which substrate binds to which aspartate-rich region is not known. Ashby et al. (1990) suggested that the DDXXD motif in L7 is the allylic binding site on the basis of sequence comparisons with prenyltransferases that utilize non-

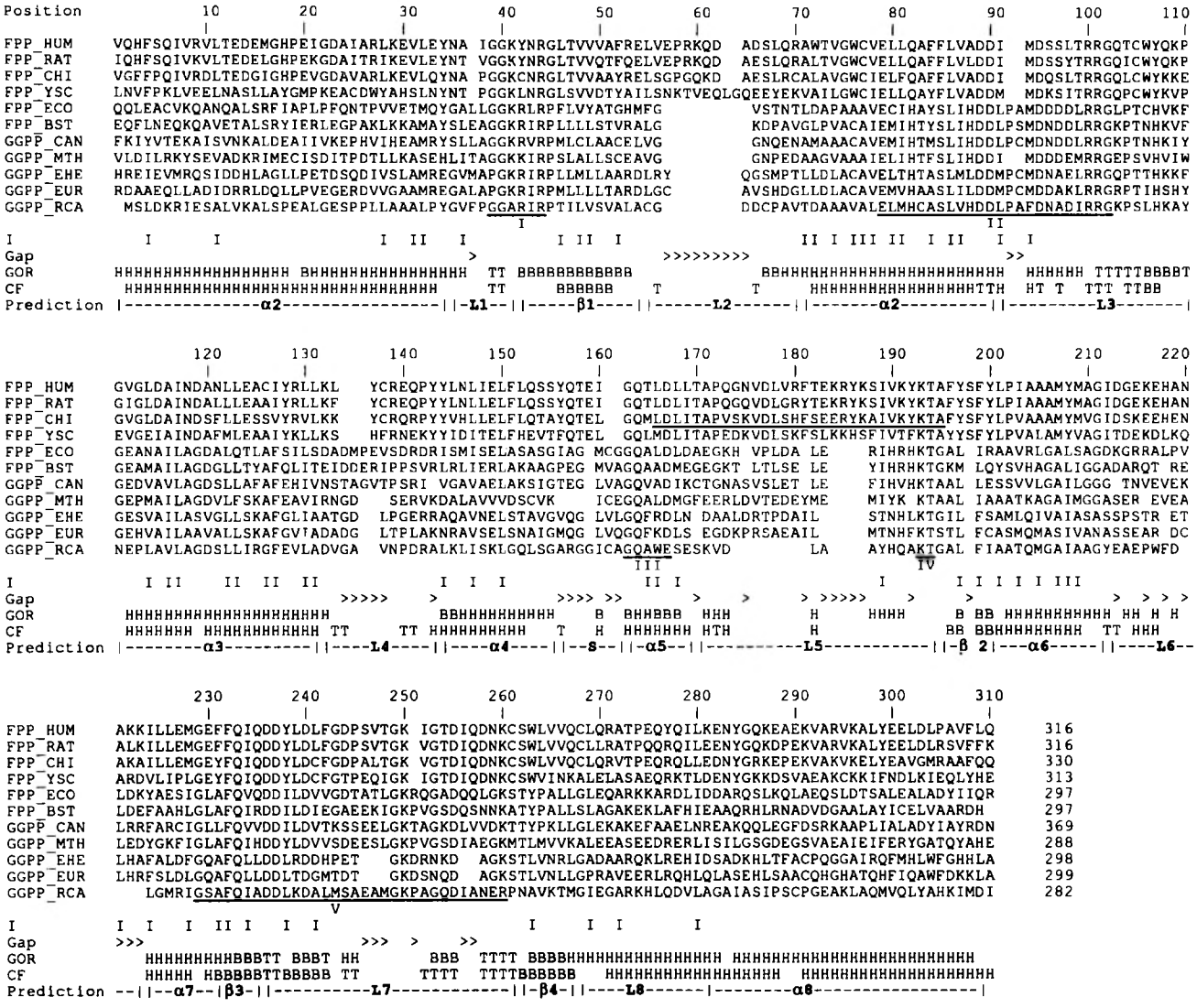

Fig. 4. A predicted consensus secondary structure based on multiple sequence alignments. The alignment of FPP synthases and GGPP synthases (not including GGPP_NCR) is the same as in Figure 2 except the N- and C-termini are omitted. The numbers on top of sequences show the alignment positions, whereas the numbers on the right are the actual amino acid numbers. The 5 conserved domains are double underlined, and the affinity labeled peptide in chicken FPP synthase is underlined. I, hydrophobic residues; CF, consensus secondary structure prediction by Chou and Fasman; GOR, consensus secondary structure prediction by GOR. The inserted gaps in the alignment are also marked by >. H and α, α-helix; B and β, β-sheet; L, loop; S, spacer; T, turn.

isoprenoid acceptors instead of IPP; however, except for the 3 aspartates, the overall sequence homologies in this region were low. A helical wheel projection of α2 and α3 indicates that a substantial portion of the total exposed surface area of these helices is hydrophobic. An alternative model for substrate binding has the diphosphate moiety in the allylic substrate interacting with the DDXX(XX)D motif in L3 instead of L7, with α2 and α3 facilitating binding of the hydrophobic isoprenoid tail through hydrophobic interactions. In this scenario, the diphosphate residue in IPP binds to the DDXXD region in L7 with the hydrophobic isopentenyl moiety in the region of the active site bounded by most of L5, which was labeled with the IPP photoaffinity analog.

All of the isoprenyl diphosphate synthases except the GGPPSases from *Erwinia* and *Rhodobacter* have charged side chains in 2 of the final 3 C-terminal residues. Amino acids containing positively charged side chains appear in the first and third positions in most of the enzymes. Site-directed mutagenesis of R350 in *Saccharomyces cerevisiae* FPPSase had little effect on the catalytic constants for the enzyme (Song & Poulter, 1994). However, fusion of a negatively charged EEF α-tubulin C-terminal epitope to the wild-type sequence reduced V_{max} 12-fold and was accompanied by a 14-fold increase in K_M for IPP. Laskovics and Poulter (1981) measured the individual kinetic constants for avian FPPSase and found that the rates of addition of substrates were substantially below the diffusion-controlled limits. These results are consistent with a conformational change in FPPSase upon binding of substrates. Thus, the C-terminus of the enzyme may form a flexible flap that helps seal the active site during catalysis.

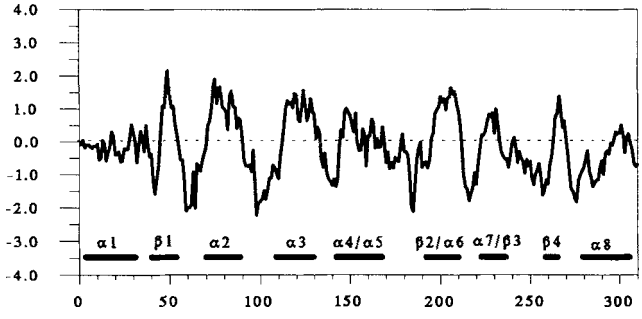

Fig. 5. An average hydropathy plot for isoprenyl diphosphate synthases. Hydropathy indices of FPP synthases and GGPP synthases (not including GGPP_NCR) were averaged at homologous positions according to the alignment shown in Figure 4. The average index is plotted along the alignment positions. Predicted α-helix and β-sheet structures are shown below the plot.

Discussion

Isoprenyl diphosphate synthases catalyze the basic chain elongation steps in the isoprenoid biosynthetic pathway. These reactions are ubiquitous in nature. Organisms contain 2 classes of isoprenyl diphosphate synthases, one for synthesis of short-chain C_{10}–C_{20} molecules and another for longer chain isoprenoids. The short-chain enzymes are further subdivided into specific enzymes for synthesis of geranyl, farnesyl, and geranylgeranyl diphosphate. The long-chain prenyltransferases are also subdivided by chain length selectivity and, in addition, specifically form either *cis* or *trans* double bonds in the newly added isoprene units. Amino acid sequences are now available for several short-chain FPPSases and GGPPSases and for 1 all-*trans* long-chain synthase. Comparisons of the primary sequences for 13 isoprenyl diphosphate synthases shown in Figure 2 revealed 5 regions containing 2–10 highly conserved amino acids. Analysis of multiply aligned sequences for the 11 FPPSases and GGPPases shown in Figure 4, in conjunction with predictions of secondary structures, indicated that these enzymes are all α-helix proteins or α/β structures dominated by α-helices.

Simple inspection of the aligned sequences suggested that individual members of the family diverged from a common ancestral isoprenyl diphosphate synthase during evolution (James et al., 1978; Bajaj & Blundell, 1984; Chothia & Lesk, 1986). A more quantitative analysis using the methods of Feng and Doolittle (1990) supported this hypothesis and provided significant insights into the pathway by which they evolved. The earliest branch was a functional segregation that separated the long-chain from the short-chain synthases, as illustrated by the large divergence between the FPPSase and the HexPPSase in yeast.

The second major branching evident in the phylogenetic tree presented in Figure 3 segregates the short-chain length prenyltransferases into 2 clusters regardless of chain length, one for eubacterial and archaebacterial proteins, and another for eukaryotic enzymes. Many organisms have distinct enzymes for synthesis of C_{10}–C_{20} isoprenyl diphosphates when these compounds serve as substrates for other enzymes. Thus, one might have anticipated a primary clustering for the short-chain enzymes according to chain length rather than kingdom. However, *M. thermoautotrophicum*, a methanogenic archaebacterium, has a single bifunctional short-chain prenyltransferase that provides both the C_{15} precursor for synthesis of squalene and the C_{20} precursor for synthesis of the distinctive isoprenoid glyceryl ether core membrane lipids found in members of the archae kingdom (Chen & Poulter, 1993). Thus, the archaebacterial enzyme may represent a primitive scenario where a single enzyme was responsible for short-chain synthesis. In this case, the fine tuning of chain length control would have evolved independently after eukaryotes and eubacteria diverged. Additional examples of eukaryotic GGPP synthases should help clarify this point. The single exception to the clustering pattern for the short-chain syntheses is the chromoplastic GGPPSase from peppers, where the gene for the enzyme may have been captured from an ancient bacterial symbiote.

It is unclear what mechanism regulates how many molecules of IPP are added to the growing isoprenoid chain by a prenyltransferase, and there appear to be no clues from the amino acid sequences shown in Figure 2. This question will not be resolved until more sequence information is available or X-ray structures are obtained for prenyltransferases with different chain-length selectivities.

Although the correlations we discovered provide important clues about the evolution of isoprenyl diphosphate synthases,

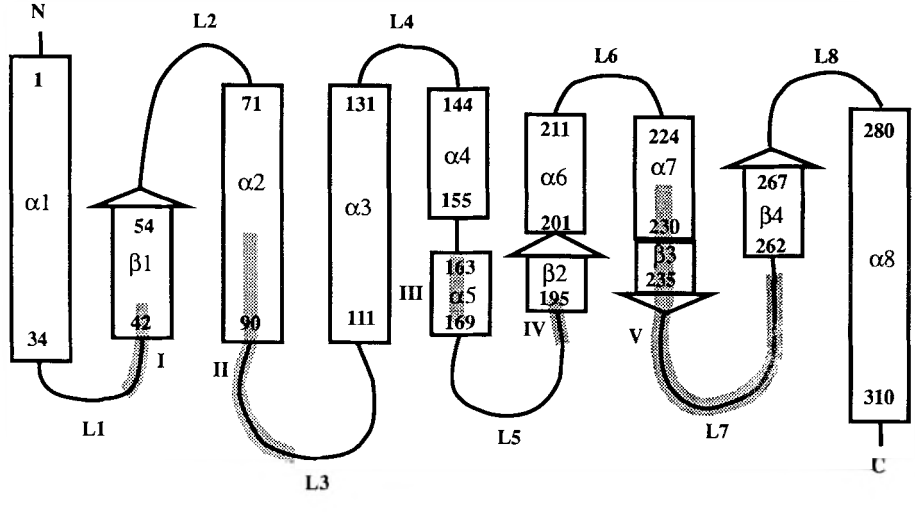

Fig. 6. The predicted average secondary structure of FPP synthases and GGPP synthases. α-Helices (α1–8) and β-sheets (β1–4) are drawn in rectangles and arrows, respectively. Loops (L1–8) are shown as curved lines. Each secondary structural unit is numbered by its position in the alignment shown in Figure 4 (see text for alternative views on α4–7 and β2–3). The 5 conserved domains (shaded) are labeled I–V. The drawing is not to scale.

the phylogenetic tree is not complete. There are no sequences yet reported for a long-chain *cis* double-bond synthase, and more examples of eukaryotic GGPP synthases are needed to confirm the groupings we propose. With the high level of activity in this area at present, these gaps should be filled in the near future.

Methods

The protein sequences of all isoprenyl diphosphate synthases except FPP_BST (Koyama et al., 1993) and FPP_CHI (Kroon et al., unpubl.) were retrieved from the Swiss-Prot data bank using GCG programs (University of Wisconsin Genetics Computer Group). The TREE program (Feng & Doolittle, 1990) was used for pairwise comparisons, to perform multiple sequence alignments, and to construct a phylogenetic tree. Refinements in the tree were made according to the protocols described by Feng and Doolittle (1990).

Average amino acid compositions of 6 FPP synthases and 6 GGPP synthases (all except GGPP_NCR) were calculated using a spreadsheet. The difference index between the average amino acid compositions of the isoprenyl diphosphate synthases and those of all-α, all-β, $\alpha + \beta$, and α/β proteins were the sum of composition differences for each amino acid, $\sum |C_{Ai} - C_{Bi}|$ (Chou, 1989; Doolittle, 1992). The smaller the difference index, the closer the comparison.

To predict an average secondary structure for FPPSases and GGPPSases, a secondary structure was calculated for each protein by the GOR procedure (Garner et al., 1978) using GARNER in PCGENE and the CF method (Chou & Fasman, 1974) using PEPTIDESTRUCTURE in GCG. The secondary structures were then arranged according to a multiple alignment truncated at both N- and C-termini. Consensus assignments of α-helix, β-sheet, or turn structures to each alignment position were determined when the assigned structural feature appeared in more than half of the aligned sequences, except at positions where gaps were inserted.

The multiple sequence alignment itself may provide information about turns and surface loops. Regions where gaps were inserted were normally considered as surface loops to accommodate insertion or deletion of a few amino acids. Additional information about structure came from hydropathy plots where regions of high hydrophobicity usually correlate with a buried β-sheet or a hydrophobic α-helix, whereas regions of high hydrophilicity usually correlate with a surface loop or a turn. An amphipathic α-helix normally occurs where there is no high or low peak in hydropathy plot. Hydropathy indices were calculated for each prenyltransferase by the method of Kyte and Doolittle (1982) using PEPTIDESTRUCTURE. These values were averaged to calculate a hydropathy index at the corresponding positions in the multiple alignment. The average indices were plotted along the alignment positions. The consensus secondary structure was predicted by combining GOR and CF structures, consideration of gaps in the alignment, and comparisons with the average hydropathy plot. The hydrophobic nature of side chains at positions containing L, I, V, M, F, W, Y, A, or C in at least 9 of 11 sequences was also indicated as I (for interior) to help visualize the hydrophobicity of secondary structure units. Helical wheel projections were constructed by HELWHEEL in PCGENE to facilitate analysis of the surface of the α-helices in the consensus structure (Shiffer & Edmundson, 1967).

Acknowledgments

We thank R.F. Doolittle for providing copies of his program and for helpful discussions. This work was supported by NIH grant GM 21328.

References

Allen CM. 1985. Purification and characterization of undecaprenyl pyrophosphate synthetase. *Methods Enzymol 110*:281–299.

Anderson MS, Yarger JG, Burck CL, Poulter CD. 1989. Farnesyl diphosphate synthetase. Molecular cloning, sequence, and expression of an essential gene from *Saccharomyces cerevisiae*. *J Biol Chem 264*:19176–19184.

Armstrong GA, Alberti M, Hearst JE. 1990. Conserved enzymes mediate the early reactions of carotenoid biosynthesis in nonphotosynthetic and photosynthetic prokaryotes. *Proc Natl Acad Sci USA 87*:9975–9979.

Armstrong GA, Alberti M, Leach F, Hearst JE. 1989. Nucleotide sequence, organization, and nature of the protein products of the carotenoid biosynthesis gene cluster of *Rhodobacter capsulatus*. *Mol Gen Genet 216*:254–268.

Ashby MN, Edwards PA. 1990. Elucidation of the deficiency in two yeast coenzyme Q mutants. *J Biol Chem 265*:13157–13164.

Ashby MN, Spear DH, Edwards PA. 1990. Prenyltransferases from yeast to man. In: Attic AD, ed. *Molecular biology of atherosclerosis*. Amsterdam: Elsevier Science Publishers. pp 27–34.

Bajaj M, Blundell T. 1984. Evolution and the tertiary structure of proteins. *Annu Rev Biophys Bioeng 13*:453–492.

Blanchard L, Karst F. 1993. Characterization of a lysine-to-glutamic acid mutation in a conservative sequence of farnesyl diphosphate synthase from *Saccharomyces cerevisiae*. *Gene 125*:185–189.

Brems DN, Bruenger E, Rilling HC. 1981. Isolation and characterization of a photoaffinity-labeled peptide from the catalytic site of prenyltransferase. *Biochemistry 20*:3711–3718.

Brems DN, Rilling HC. 1979. Photoaffinity labeling of the catalytic site of prenyltransferase. *Biochemistry 18*:860–864.

Cane DE. 1981. Biosynthesis of sesquiterpenes. In: Porter JW, Spurgeon SL, eds. *Biosynthesis of isoprenoid compounds, vol 1*. New York: John Wiley & Sons. pp 283–374.

Carattoli A, Romano N, Ballario P, Morelli G, Macino G. 1991. The *Neurospora crassa* carotenoid biosynthetic gene (Albino 3) reveals highly conserved regions among prenyltransferases. *J Biol Chem 266*:5854–5859.

Chen A, Poulter CD. 1993. Purification and characterization of farnesyl diphosphate/geranylgeranyl diphosphate synthase. A thermostable bifunctional enzyme from *Methanobacterium thermoautotrophicum*. *J Biol Chem 268*:11002–11007.

Chothia C, Lesk AM. 1986. The relation between the divergence of sequence and structure in proteins. *EMBO J 5*:823–826.

Chou PY. 1989. Prediction of protein structural classes from amino acid compositions. In: Fasman GD, ed. *Prediction of protein structure and the principles of protein conformation*. New York: Plenum Press. pp 549–586.

Chou PY, Fasman GD. 1974. Prediction of protein conformation. *Biochemistry 13*:222–224.

Clarke CF, Tanaka RD, Svenson K, Wamsley M, Fogelman AM, Edwards PA. 1987. Molecular cloning and sequence of a cholesterol-repressible enzyme related to prenyltransferase in the isoprene biosynthetic pathway. *Mol Cell Biol 7*:3138–3146.

Clarke S. 1992. Protein isoprenylation and methylation at carboxyl-terminal cysteine residues. *Annu Rev Biochem 61*:355–386.

Crawford IP, Niermann T, Kirschner K. 1987. Prediction of secondary structure by evolutionary comparison: Application to the α subunit of tryptophan synthase. *Proteins Struct Funct Genet 2*:118–129.

Croteau R. 1981. Biosynthesis of monoterpenes. In: Porter JW, Spurgeon SL, eds. *Biosynthesis of isoprenoid compounds, vol 1*. New York: John Wiley & Sons. pp 225–282.

Doolittle RF. 1992. A detailed consideration of a principal domain of vertebrate fibrinogen and its relatives. *Protein Sci 1*:1563–1577.

Feng DF, Doolittle RF. 1987. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J Mol Evol 25*:351–360.

Feng DF, Doolittle RF. 1990. Progressive alignment and phylogenetic tree construction of protein sequences. *Methods Enzymol 183*:375–387.

Fujisaki S, Hara H, Nishimura Y, Horiuchi K, Nishino T. 1990. Cloning and nucleotide sequence of *ispA* gene responsible for farnesyl diphosphate synthase activity in *Escherichia coli*. *J Biochem 108*:995–1000.

Garner J, Osguthorpe DJ, Robson B. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J Mol Biol 120*:97–120.

James MNG, Delbaere LTJ, Brayer GD. 1978. Amino acid sequence align-

ment of bacterial and mammalian pancreatic serine proteases based on topological equivalences. *Can J Biochem 56*:396–402.

Joly A, Edwards PA. 1993. Effect of site-directed mutagenesis of conserved aspartate and arginine residues upon farnesyl diphosphate synthase activity. *J Biol Chem 268*:26983–26989.

Koyama T, Obata S, Osabe M, Takeshita A, Yokoyama K, Uchida M, Nishino T, Ogura K. 1993. Thermostable farnesyl diphosphate synthase of *Bacillus stearothermophilus*: Molecular cloning, sequence determination, over production, and purification. *J Biochem 113*:355–363.

Kuntz M, Roemer S, Suire C, Hugueney P, Well JH, Schantz R, Camara B. 1992. Identification of a cDNA for the plastid-located geranylgeranyl pyrophosphate synthase from *Capsicum annuum*: Correlative increase in enzyme activity and transcript level during fruit ripening. *Plant J 2*:25–34.

Kyte J, Doolittle RF. 1982. A simple method for displaying the hydropathic character of a protein. *J Mol Biol 157*:105–132.

Laskovics FM, Poulter CD. 1981. Prenyltransferase: Determination of the binding mechanism and individual kinetic constants for farnesyl pyrophosphate synthetase by rapid quench and isotope partitioning experiments. *Biochemistry 20*:1893–1901.

Marrero PF, Poulter CD, Edwards PA. 1992. Effects of site-directed mutagenesis of the highly conserved aspartate residues in domain II of farnesyl diphosphate synthase activity. *J Biol Chem 267*:21873–21878.

Math SK, Hearst JE, Poulter CD. 1992. The *crtE* gene in *Erwiniaherbicola* encodes geranylgeranyl diphosphate synthase. *Proc Natl Acad Sci USA 89*:6761–6764.

Matsuoka S, Sagami H, Kurisaki A, Ogura K. 1991. Variable product specificity of microsomal dehydrodolichyl diphosphate synthase from rat. *J Biol Chem 266*:3464–3468.

Misawa N, Nakagawa M, Kobayashi K, Yamano S, Izawa Y, Nakamura K, Harashima K. 1990. Elucidation of the *Erwinia uredovora* carotenoid biosynthetic pathway by functional analysis of gene products expressed in *Escherichia coli*. *J Bacteriol 172*:6704–6712.

Myers EW, Miller W. 1988. Optimal alignments in linear space. *Comput Applic Biosci 4*:11–17.

Poulter CD, Rilling HC. 1978. The prenyltransferase reaction. Enzymatic and mechanistic studies of the 1′-4 coupling reaction in the terpene biosynthetic pathway. *Acc Chem Res 11*:307–313.

Poulter CD, Rilling HC. 1981a. Conversion of farnesyl pyrophosphate to squalene. In: Porter JW, Spurgeon SL, eds. *Biosynthesis of isoprenoid compounds, vol 1*. New York: John Wiley & Sons. pp 225–282.

Poulter CD, Rilling HC. 1981b. Prenyltransferases and isomerase. In: Porter JW, Spurgeon SL, eds. *Biosynthesis of isoprenoid compounds, vol 1*. New York: John Wiley & Sons. pp 161–224.

Schiffer M, Edmundson AB. 1967. Use of helical wheels to represent the structures of proteins and to identify segments with helical potential. *Biophys J 7*:121–135.

Sheares BT, White SS, Molowa DT, Chan K, Ding VDH, Kroon PA, Bostedor RG, Karkas JD. 1989. Cloning, analysis, and bacterial expression of human farnesyl pyrophosphate synthetase and its regulation in Hep G2 cell. *Biochemistry 28*:8129–8135.

Song LS, Poulter CD. 1994. Yeast farnesyl diphosphate synthase. Site-directed mutagenesis of residues in highly conserved prenyltransferase domain I and domain II. *Proc Natl Acad Sci USA*. Forthcoming.

Spurgeon SL, Porter JW. 1981. Biosynthesis of carotenoids. In: Porter JW, Spurgeon SL, eds. *Biosynthesis of isoprenoid compounds, vol 2*. New York: John Wiley & Sons. pp 1–122.

West CA. 1981. Biosynthesis of diterpenes. In: Porter JW, Spurgeon SL, eds. *Biosynthesis of isoprenoid compounds, vol 1*. New York: John Wiley & Sons. pp 375–411.